ORIGINAL RESEARCH

Nomogram for the Prediction of Biochemical Incomplete Response in Papillary Thyroid Cancer Patients

Seung Taek Lim[1]
Ye Won Jeon[1]
Hongki Gwak[1]
Ja Seong Bae[2]
Young Jin Suh[1]

[1]Division of Breast & Thyroid Surgical Oncology, Department of Surgery, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Suwon, Korea; [2]Division of Thyroid Surgical Oncology, Department of Surgery, Seoul St. Mary's Hospital, Catholic University of Korea College of Medicine, Seoul, Korea

Purpose: To develop a nomogram for predicting biochemical incomplete response (BIR) in the dynamic risk stratification (DRS) of papillary thyroid carcinoma (PTC) patients without structural recurrence, and to investigate its validity.

Patients and Methods: Overall, 1705 (1005 and 700 in the training and validation cohorts, respectively) PTC patients treated with total thyroidectomy without structural recurrence were included. multivariate logistic regression analyses were performed to determine the significant predictors of BIR in the training cohort. A nomogram was subsequently constructed for BIR risk prediction. Assessments for the predictive accuracy, discrimination, and calibration of the nomogram were performed. Subsequently, internal and external validations were conducted.

Results: In the multivariate analysis, age, sex, lymph node metastasis site, extrathyroidal extension, and lymphovascular invasion showed significant predictive value; using these predictive factors and tumor size, a nomogram for BIR risk prediction was constructed. In the training cohort, the nomogram showed good predictive performance and discrimination in the receiver operating characteristic (ROC) curve analysis, with an area under the curve (AUC) of 0.765. In internal validation, the bootstrap-corrected AUC was 0.76. The calibration plot showed good agreement between the predicted and actual observation. The Hosmer–Lemeshow (HL) test did not suggest a lack of fit (p=0.1613). In the external validation, the AUC was 0.828 in the ROC curve analysis; the calibration plot showed good quality, and the HL test did not suggest a lack of fit (p=0.2161).

Conclusion: The constructed nomogram may effectively predict the risk of BIR in DRS in PTC patients without structural recurrence.

Level of Evidence: Level 4.

Keywords: papillary thyroid carcinoma, dynamic risk stratification, nomogram, biochemical incomplete response

Introduction

Thyroid cancer is a common endocrine malignancy, with papillary thyroid carcinoma (PTC) being the most frequently observed histologic subtype.[1–4] Previous studies have shown favorable prognoses associated with PTC;[5,6] however, the incidence of locoregional recurrence is significant in PTC patients during the follow-up period. The risk of recurrence was initially predicted using the staging system, which includes patients' disease characteristics at diagnosis.[7–9]

In the 2015 American Thyroid Association (ATA) guideline, the dynamic risk stratification (DRS) system was proposed to evaluate individual responses to initial

Correspondence: Young Jin Suh
Division of Breast & Thyroid Surgical Oncology, Department of Surgery, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Suwon, Korea
Tel +82-31-249-8169
Fax +82-31-247-5347
Email yjsuh@catholic.ac.kr

therapy in PTC patients.[10] This tool is equivalent to the initially used staging system, as recurrence risk in the initial staging system can be re-stratified based on the patients' response to initial therapy, allowing for the guidance of follow-up management strategies.

In the DRS system, patients are classified into four response categories according to their imaging and laboratory test results: excellent response (ER), indeterminate response (IR), biochemical incomplete response (BIR), and structural incomplete response (SIR). Although the disease spontaneously evolves to "no evidence of disease" with additional therapy in a majority of these patients with BIR, BIR may be clinically significant, as 19–27% of patients experience persistently abnormal high thyroglobulin (Tg) levels and 8–17% of them experience structural recurrence in the 5–10-year follow-up period after initial therapy.[10] Although the management guidelines for patients with SIR suggest the extirpation of recurrence lesions, there are no distinct guidelines for the management of patients with BIR.

Accordingly, clinician-related variations exist in management strategies for patients showing a BIR, including active surveillance with frequent laboratory and imaging studies, additional radioactive iodine (RAI) ablation therapy, or sustained thyroid-stimulating hormone (TSH) suppression therapy that may not be necessary.[11] Moreover, patients with BIR may be exposed to unwarranted imaging studies to identify possible recurrence lesions, leading to unnecessary biopsy or surgery and eventually increasing morbidity and worsening patients' quality of life. However, previous studies have reported only individual risk factors for BIR; there is no comprehensive prediction model including risk factors for the estimation of BIR risk. Therefore, the development of a reliable prediction model for the estimation of BIR risk on DRS could aid in the prediction of prognoses and provide reliable information on the clinical course and surveillance recommendations for PTC patients.

Therefore, in this study, we aimed to develop a nomogram for the prediction of BIR in the DRS of PTC patients without structural recurrence, and to investigate its validity through internal and external validations.

Patients and Methods
Study Cohorts
We retrospectively reviewed the patients' electronic medical records for clinicopathologic data and classified them into the training and validation cohorts based on two independent hospital datasets. In the training cohort, the data of 1104 PTC patients aged >18 years who were treated with total thyroidectomy (TT) in St. Vincent's Hospital between January 2012 and December 2017 were examined. In the validation cohort, the data of 776 PTC patients aged >18 years who were treated with TT in Seoul St. Mary's Hospital between May 2013 and December 2015 were collected.

Ethical Statement
This study was approved by the Institutional Review Board (IRB) of St. Vincent's Hospital (IRB No. XC20RADE0141). All procedures involving human participants were conducted in accordance with the ethical standards of the Institutional Research Committee and the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All participants provided written informed consent for storing medical information in the hospital database and use this information for research purposes.

Data Collection and Patient Inclusion
Data on the baseline clinicopathologic characteristics outlined in Table 1 were obtained. Central lymph node dissection (CLND) was performed for prophylaxis or treatment at the surgeon's discretion. When CLND was performed prophylactically, the ipsilateral or bilateral central neck compartment in the same side as the PTC was dissected, even in the absence of clinical suspicions of lymph node (LN) metastasis. When CLND was performed therapeutically, the ipsilateral or bilateral central neck compartment was dissected only when metastatic LNs were preoperatively identified, as confirmed by biopsy or suspicions on imaging study. Modified radical neck dissection (MRND) was performed therapeutically only when the presence of metastatic LNs in the lateral neck compartment (levels II, III, IV, V) was confirmed by biopsy or preoperative imaging studies. RAI ablation after surgery was performed at the clinician's or multidisciplinary team's discretion.

All patients were clinically evaluated and underwent laboratory tests, including those for the serum free thyroxine, triiodothyronine, TSH, Tg, and serum antithyroglobulin antibody (anti-TgAb) levels, every 3–6 months within 2 years of initial treatment. Imaging studies, including neck ultrasonography, chest radiography, and neck computed tomography (CT), were routinely performed. Additional imaging studies, including chest CT, diagnostic whole-body iodine scan, or positron emission tomography with fluorodeoxyglucose/CT, were performed if needed. The serum TSH levels were considered stimulated at >30 IU/mL following 4 weeks of levothyroxine hormone

Table 1 Baseline Characteristics of Patients

	All Patients			Training Cohort			Validation Cohort		
	Training Cohort (N=1005)	Validation Cohort (N=700)	p-value	Non-BIR (N=926)	BIR (N=79)	p-value	Non-BIR (N=659)	BIR (N=41)	p-value
Age, years									
<55	618 (61.5%)	469 (67.0%)	0.02	573 (61.9%)	45 (57.0%)	0.389	440 (66.8%)	29 (70.7%)	0.6
≥55	387 (38.5%)	231 (33.0%)		353 (38.1%)	34 (43.0%)		219 (33.2%)	12 (29.3%)	
Sex									
Female	855 (85.1%)	520 (74.3%)	<0.001	798 (86.2%)	57 (72.2%)	0.001	495 (75.1%)	25 (61.0%)	0.044
Male	150 (14.9%)	180 (25.7%)		128 (13.8%)	22 (27.8%)		164 (24.9%)	16 (39.0%)	
Tumor size									
<2 cm	932 (92.7%)	616 (88.0%)	0.001	865 (93.4%)	67 (84.8%)	0.005	592 (89.8%)	24 (58.5%)	<0.001
≥2 cm	73 (7.3%)	84 (12.0%)		61 (6.6%)	12 (15.2%)		67 (10.2%)	17 (41.5%)	
Tumor number									
1	565 (56.2%)	287 (41.0%)	<0.001	527 (56.9%)	38 (48.1%)	0.165	267 (40.5%)	20 (48.8%)	0.573
2	249 (24.8%)	209 (29.9%)		229 (24.7%)	20 (25.3%)		198 (30.0%)	11 (26.8%)	
More than 3	191 (19.0%)	204 (29.1%)		170 (18.4%)	21 (26.6%)		194 (29.5%)	10 (24.4%)	
Histologic subtype									
Classic	899 (89.4%)	550 (78.6%)	<0.001	827 (89.3%)	72 (91.1%)	0.951	517 (78.5%)	33 (80.5%)	0.574
Follicular variant	66 (6.6%)	82 (11.7%)		62 (6.7%)	4 (5.1%)		79 (12.0%)	3 (7.3%)	
Tall cell	28 (2.8%)	56 (8.0%)		26 (2.8%)	2 (2.5%)		51 (7.7%)	5 (12.2%)	
Warthin-like	12 (1.2%)	12 (1.7%)		11 (1.2%)	1 (1.3%)		12 (1.8%)	0 (0%)	
Extrathyroidal extension									
No	506 (50.3%)	262 (37.4%)	<0.001	483 (52.2%)	23 (29.1%)	<0.001	257 (39.0%)	5 (12.2%)	<0.001
Microscopic	438 (43.6%)	375 (53.6%)		391 (42.2%)	47 (59.5%)		352 (53.4%)	23 (56.1%)	
Gross	61 (6.1%)	63 (9.0%)		52 (5.6%)	9 (11.4%)		50 (7.6%)	13 (31.7%)	
Lymphovascular invasion									
No	747 (74.3%)	412 (58.9%)	<0.001	715 (77.2%)	32 (40.5%)	<0.001	406 (61.6%)	6 (14.6%)	<0.001
Yes	258 (25.7%)	288 (41.1%)		211 (22.8%)	47 (59.5%)		253 (38.4%)	35 (85.4%)	
Multifocality									
No	567 (56.4%)	287 (41.0%)	<0.001	529 (57.1%)	38 (48.1%)	0.12	267 (40.5%)	20 (48.8%)	0.297
Yes	438 (43.6%)	413 (59.0%)		397 (42.9%)	41 (51.9%)		392 (59.5%)	21 (51.2%)	
Bilaterality									
No	697 (69.4%)	412 (58.9%)	<0.001	648 (70.0%)	49 (62.0%)	0.141	388 (58.9%)	24 (58.5%)	0.966
Yes	308 (30.6%)	288 (41.1%)		278 (30.0%)	30 (38.0%)		271 (41.1%)	17 (41.5%)	
Extent of LN excision									
No	275 (27.3%)	9 (1.3%)	<0.001	261 (28.2%)	14 (17.7%)	<0.001	9 (1.4%)	0 (0%)	<0.001
CLND	660 (65.7%)	522 (74.6%)		612 (66.1%)	48 (60.8%)		502 (76.2%)	20 (48.8%)	
MRND	70 (7.0%)	169 (24.1%)		53 (5.7%)	17 (21.5%)		148 (22.4%)	21 (51.2%)	
Site of LN metastasis									
No	696 (69.3%)	249 (35.6%)	<0.001	668 (72.1%)	28 (35.5%)	<0.001	248 (37.6%)	1 (2.4%)	<0.001
Central	243 (24.2%)	320 (45.7%)		209 (22.6%)	34 (43.0%)		299 (45.4%)	21 (51.2%)	
Lateral	66 (6.6%)	131 (18.7%)		49 (5.3%)	17 (21.5%)		112 (17.0%)	19 (46.4%)	
Number of metastatic LNs	1.27±3.43	3.77±5.47	<0.001	0.97±2.57	4.89±7.67	<0.001	3.33±4.92	10.68±8.46	<0.001
Number of excised LNs	5.08±8.00	20.15±21.13	<0.001	4.73±7.44	9.32±12.18	0.001	19.28±20.43	34.21±26.84	0.001
Metastatic/excised LN ratio	0.15±0.28	0.18±0.22	0.011	0.13±0.26	0.40±0.37	<0.001	0.17±0.21	0.39±0.25	<0.001

(Continued)

Table 1 (Continued).

	All Patients			Training Cohort			Validation Cohort		
	Training Cohort (N=1005)	**Validation Cohort (N=700)**	**p-value**	**Non-BIR (N=926)**	**BIR (N=79)**	**p-value**	**Non-BIR (N=659)**	**BIR (N=41)**	**p-value**
RAI ablation treatment			<0.001			<0.001			<0.001
No	649 (64.6%)	220 (31.4%)		626 (67.6%)	23 (29.1%)		218 (33.1%)	2 (4.9%)	
Yes	356 (35.4%)	480 (68.6%)		300 (32.4%)	56 (70.9%)		441 (66.9%)	39 (95.1%)	
Interval between surgery and RAI ablation (months)	3.62±2.79	2.79±0.71	<0.001	3.53±1.86	4.12±5.54	0.431	2.81±0.72	2.64±0.62	0.151
RAI ablation treatment dose			<0.01			0.718			1
100 mci	129 (36.2%)	480 (99.8%)		109 (36.5%)	19 (33.9%)		441 (99.8%)	39 (100%)	
150 mci	227 (62.8%)	1 (0.2%)		190 (63.5%)	37 (66.1%)		1 (0.2%)	0 (0%)	
Interval between initial treatment and DRS (months)	12.15±3.81	12.14±3.25	0.949	12.15±3.78	12.17±4.16	0.963	12.27±3.16	10.07±3.90	<0.001
Tg level in the DRS (ng/mL)	2.29±15.59	0.79±4.57	0.004	0.43±1.22	24.09±50.90	<0.001	0.11±0.25	11.75±15.30	<0.001
TSH level in the DRS (IU/mL)	12.07±27.39	1.88±10.78	<0.001	10.67±24.88	28.56±44.93	0.001	1.84±10.78	2.45±11.00	0.727
TSH stimulation in the DRS			<0.001			<0.001			0.518
No	850 (84.6%)	688 (98.3%)		794 (85.7%)	56 (70.9%)		648 (98.3%)	40 (97.6%)	
Yes	155 (15.4%)	12 (1.7%)		132 (14.3%)	23 (29.1%)		11 (1.7%)	1 (2.4%)	
Result of DRS			<0.001			<0.001			<0.001
ER	673 (67.0%)	602 (86.0%)		673 (72.7%)	0 (0%)		602 (91.4%)	0 (0%)	
IR	253 (25.2%)	57 (8.1%)		253 (27.3%)	0 (0%)		57 (8.6%)	0 (0%)	
BIR	79 (7.9%)	41 (5.9%)		0 (0%)	79 (100%)		0 (0%)	41 (100%)	

Abbreviations: BIR, biochemical incomplete response; CLND, central lymph node dissection; MRND, modified radical neck dissection; RAI, radioactive iodine; Tg, thyroglobulin; TSH, thyroid stimulating hormone; DRS, dynamic risk stratification; ER, excellent response; IR, indeterminate response, LN, lymph node.

withdrawal, whereas the serum anti-TgAb levels were considered abnormal at ≥40 IU/mL. The serum Tg and anti-TgAb levels were measured using the Siemens Immulite® 2000 thyroglobulin and antithyroglobulin assay (Siemens Healthineers, Erlangen, Germany), with the same method employed in the two hospitals.

Patients were excluded because of the following: presence of distant metastasis at diagnosis, absence of adequate clinicopathologic data for analysis, abnormal anti-TgAb levels at diagnosis or in the follow-up period, and experience of structural recurrence before DRS. In accordance with the exclusion criteria, three patients with distant metastasis at diagnosis, 33 without adequate clinicopathologic data for the analysis, 42 with abnormal anti-TgAb levels at diagnosis or in the follow-up period, and 21 with structural recurrence before DRS were excluded in the training cohort. In the validation cohort, 59 patients without adequate clinicopathologic data, 15 with abnormal anti-TgAb levels at diagnosis or in the follow-up period, and two with structural recurrence before DRS were

excluded. Finally, 1005 and 700 patients were assigned to the training and validation cohorts, respectively.

The DRS results were evaluated according to the stimulated or unstimulated serum Tg levels, as obtained at 6–24 months after the completion of the initial therapy. Among patients who underwent TT with RAI ablation, the DRS results were determined as follows: ER, unstimulated Tg level <0.2 ng/mL or stimulated Tg level <1 ng/mL and negative findings on imaging study; IR, unstimulated Tg level 0.2–1 ng/mL or stimulated Tg level 1–10 ng/mL and negative findings on imaging study; and BIR, unstimulated Tg level >1 ng/mL or stimulated Tg level >10 ng/mL and negative findings on imaging study. Among patients who underwent TT without RAI ablation, DRS results were determined as follows: ER, unstimulated Tg level <0.2 ng/mL or stimulated Tg level <2 ng/mL and negative findings on imaging study; IR, unstimulated Tg level 0.2–5 ng/mL or stimulated Tg level 2–10 ng/mL and negative findings on imaging study; and BIR, unstimulated Tg level >5 ng/mL or

stimulated Tg level >10 ng/mL and negative findings on imaging study.

Statistical Analyses

Descriptive characteristics are summarized as numbers and percentages for categorical variables and as means±standard deviations for continuous variables. For comparing the clinicopathologic characteristics between the groups, the Student's *t*-test or Mann–Whitney *U*-test was used for continuous variables, and the chi-squared or Fisher's exact test was used for categorical variables to assess the differences in the distribution of the variables.

Regarding the construction of the nomogram, data on the training cohort were used for nomogram model development. First, univariate and multivariate logistic regression analyses with the backward stepwise procedure, based on the Akaike information criterion, were performed to determine significant predictors of BIR among the variables in the training cohort. Subsequently, these significant predictors were selected for the development and construction of a nomogram for BIR prediction. To construct the nomogram, each regression beta-coefficient of variables in the multivariate logistic regression was proportionally converted to a 0–100-point scale. The variable with the highest beta-coefficient value was assigned 100 points, and the others were assigned 0–100 points according to the beta-coefficient value relative to the highest one. After the assignment of scores to each variable, the individual scores for each patient in the training and validation cohorts were calculated, which were subsequently used to estimate the individual BIR risk. To assess the predictive accuracy and discrimination of the nomogram in the training cohort, receiver operating characteristic (ROC) curves with areas under the curve (AUC) were calculated using the individual scores. Next, the internal validity of the prediction model was tested with 1000-bootstrap resampling to reduce the effect of overfitting bias. The degree of calibration was assessed with calibration plots (1000-bootstrap resampling) and the Hosmer–Lemeshow (HL) goodness-of-fit test. In the HL goodness-of-fit test, the level of model calibration assessment was considered adequate at values of p>0.05. For external validation, the nomogram performance was assessed in the validation cohort using the same methods as those employed in the training cohort.

All statistical analyses were performed using SPSS Statistics for Windows, version 12.0 (SPSS Inc., Chicago, IL, USA) or R version 4.0 (R Foundation for Statistical Computing, Vienna, Austria). Odds ratios (ORs) were presented with their 95% confidence intervals (CIs). All statistical tests were two-sided, and the significance level was set at p<0.05.

Results

Baseline Characteristics

Overall, 1705 (1005, training cohort; 700, validation cohort) patients were included. In the training cohort, 79 patients (7.86%) had BIR on DRS and 926 (92.14%) had no BIR. In the validation cohort, 41 patients (5.85%) had BIR on DRS and 659 (94.15%) had no BIR. The clinicopathologic data on the study cohort are presented in Table 1.

In the training cohort, the BIR group had significantly more male patients; a larger tumor size; higher rates of extrathyroidal extension (ETE), lymphovascular invasion (LVI), RAI, and MRND; and increased incidence of central and lateral LN metastases, number of metastatic LNs, number of excised LNs, and metastatic/excised LN ratio than the non-BIR group. The same findings were observed in the validation cohort; in addition, the BIR group showed a shorter interval between the initial treatment and DRS, and increased Tg levels in the DRS.

Between the training cohort and validation cohort, significant differences were observed in the following parameters: age, sex, tumor size, tumor number, histologic subtype, ETE, LVI, multifocality, bilaterality, extent of LN excision, site of LN metastasis, number of metastatic LNs, number of excised LNs, metastatic/excised LN ratio, RAI ablation treatment, interval between surgery and RAI ablation, RAI ablation treatment dose, stimulated Tg and TSH levels in the DRS, and DRS result.

Determining the Predictors of BIR and Nomogram Construction

To determine the predictors of BIR, univariate and multivariate logistic regression analyses were performed (Table 2). The multivariate analysis demonstrated that age ≥55 years (OR: 1.686, 95% CI: 1.018–2.794, p=0.042), male sex (OR: 1.851, 95% CI: 1.037–3.303, p=0.037), site of LN metastasis (central: OR: 2.316, 95% CI: 1.120–4.791, p=0.024; lateral: OR: 3.470, 95% CI: 1.416–8.500, p=0.007), presence of ETE (microscopic: OR: 1.804, 95% CI: 1.049–3.105, p=0.033; gross: OR: 2.547, 95% CI: 1.059–6.127, p=0.037), and presence of LVI (OR: 2.146, 95% CI: 1.076–4.282, p=0.03) were significantly associated with BIR.

In addition to the abovementioned five significant predictors, the tumor size was included to construct the nomogram for the prediction of BIR risk, as it is a well-known prognostic

Table 2 Univariate and Multivariate Logistic Regression Analyses for the Prediction of BIR Risk Factors in the Training Cohort

Variables		Univariate		Multivariate[a]		
		OR (95% CI)	p-value	Beta	OR (95% CI)	p-value
				−3.917[b]		
Age, years	<55	I (reference)			I (reference)	
	≥55	1.694 (1.018–2.819)	0.043	0.523	1.686 (1.018–2.794)	0.042
Sex	Female	I (reference)			I (reference)	
	Male	1.829 (1.018–3.288)	0.044	0.616	1.851 (1.037–3.303)	0.037
Histologic subtype	Classic	I (reference)		–	–	–
	Follicular variant	0.953 (0.322–2.820)	0.93			
	Tall cell	0.388 (0.084–1.796)	0.226			
	Warthin-like	1.025 (0.120–8.735)	0.982			
Tumor size	<2 cm	I (reference)			I (reference)	–
	≥2 cm	1.306 (0.617–2.767)	0.486	0.236	1.266 (0.602–2.661)	0.534
Site of LN metastasis	No	I (reference)			I (reference)	
	Central	2.187 (1.053–4.540)	0.036	0.84	2.316 (1.120–4.791)	0.024
	Lateral	3.296 (1.324–8.206)	0.01	1.244	3.470 (1.416–8.500)	0.007
Extrathyroidal extension	No	I (reference)			I (reference)	
	Microscopic	1.753 (1.010–3.041)	0.046	0.59	1.804 (1.049–3.105)	0.033
	Gross	2.576 (1.056–6.284)	0.037	0.935	2.547 (1.059–6.127)	0.037
Lymphovascular invasion	No	I (reference)			I (reference)	
	Yes	2.270 (1.131–4.554)	0.021	0.764	2.146 (1.076–4.282)	0.03
Multifocality	No	I (reference)		–	–	–
	Yes	1.056 (0.512–2.178)	0.882			
Bilaterality	No	I (reference)		–	–	–
	Yes	1.022 (0.483–2.161)	0.955			

Notes: [a]Adjusted for age, sex, histologic subtype, size, site of lymph node metastasis, ETE, LVI, multifocality, and bilaterality. [b]Intercept.
Abbreviations: BIR, biochemical incomplete response; OR, odds ratio; CI, confidence interval; LN, lymph node; ETE, extrathyroidal extension; LVI, lymphovascular invasion.

factor for PTC, although statistical significance was not achieved in our multivariate analysis. Finally, six predictors (age, sex, tumor size, site of LN metastasis, ETE, and LVI) were used to construct a BIR risk prediction nomogram (Figure 1).

## Internal and External Validation of the Nomogram for BIR Prediction

In the training cohort, the ROC curve analysis showed good predictive performance and discrimination, with an AUC of 0.765 (95% CI: 0.712–0.818) (Figure 2A). After bootstrapping for internal validation, the bootstrap-corrected AUC of the prediction model was 0.76 (95% CI: 0.709–0.816) with 1000 resampling. The calibration plot showed good agreement between the nomogram's

prediction and the actual observation (Figure 2B), and the HL test did not suggest a lack of fit (p=0.1613).

Subsequently, to overcome the limitations of internal validation, we conducted an external validation to determine the generalizability of the prediction model using a validation cohort. In the external validation performed in the validation cohort, the AUC was 0.828 (95% CI: 0.768–0.889) in the ROC curve analysis, and the calibration plot showed a good calibration level (Figure 3A and B). The HL test did not suggest a lack of fit (p=0.2161). Although the training and validation cohorts had different baseline characteristics, the performance of the model remained good.

Discussion

In this study, we developed and validated a nomogram for BIR risk prediction on DRS in PTC patients without
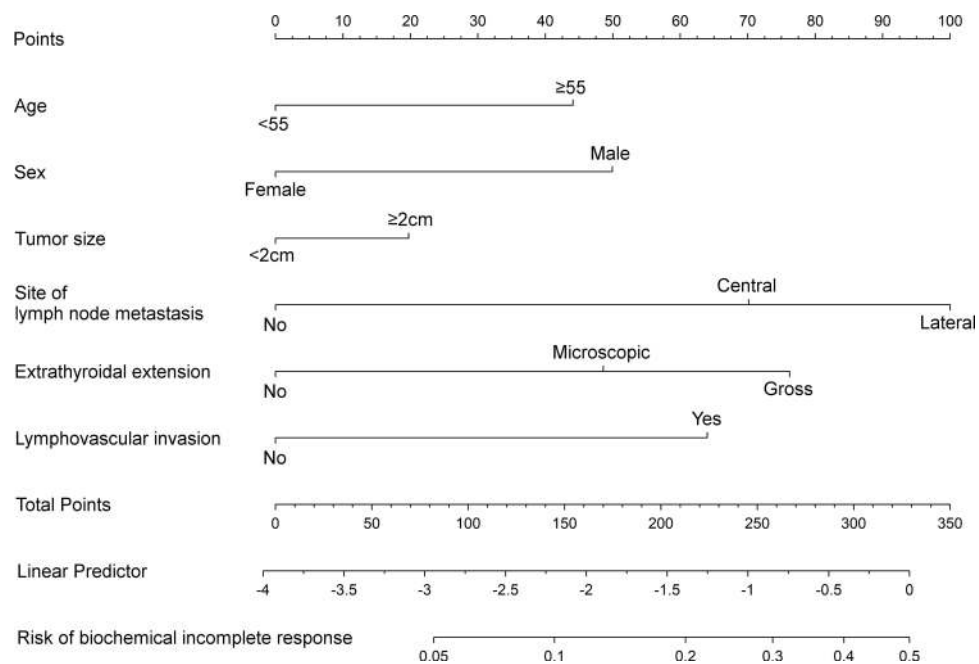
Figure 1 Nomogram for biochemical incomplete response risk prediction in papillary thyroid cancer patients.
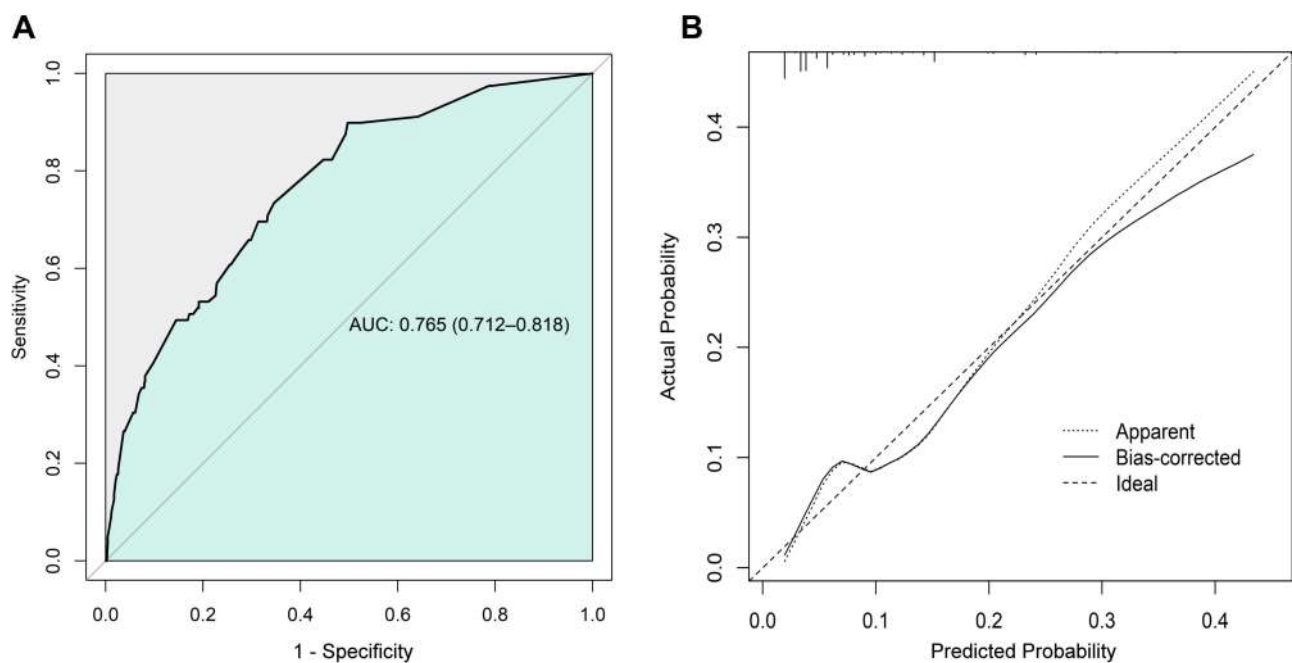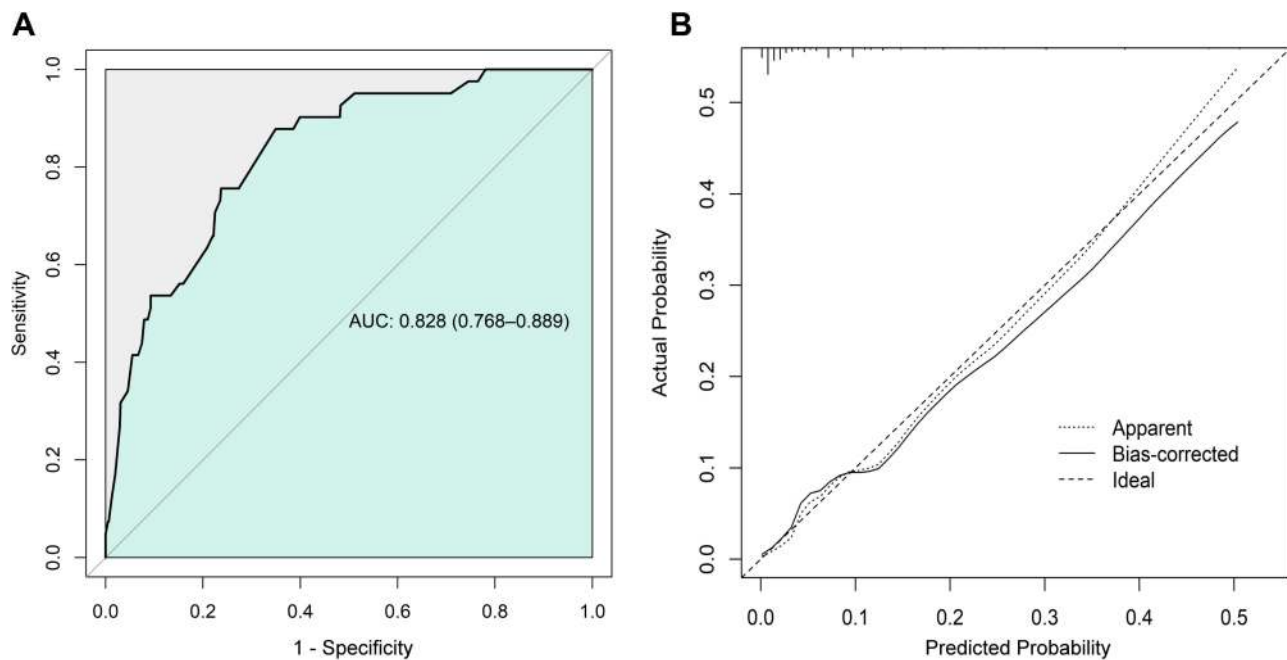

Figure 2 (**A**) Receiver operating characteristic curve analysis of the nomogram in the training cohort. (**B**) Calibration plot of the nomogram in the training cohort with bootstrapping.
Abbreviation: AUC, area under the curve.

structural recurrence using the following predictors in the multivariate analysis: age $\geq$ 55 years, male sex, central and lateral LN metastasis, ETE, and LVI. Additionally, the nomogram was successfully validated through internal and external validations using an independent cohort dataset. The clinical and pathologic risk factors in our model could be easily corrected, minimizing the degree of variation in the collection of patients'

Figure 3 (**A**) Receiver operating characteristic curve analysis of the nomogram in the validation cohort. (**B**) Calibration plot of the nomogram in the validation cohort with bootstrapping.
Abbreviation: AUC, area under the curve.

data and improving the general applicability of the prediction model.

Per the ATA guideline, a BIR on DRS is observed in 11–19%, 21–22%, and 16–18% of low-risk, intermediate-risk, and high-risk patients, respectively.[12,13] In previous studies, the definition of BIR was based on several factors, such as whether the patient underwent RAI ablation treatment or whether Tg production was stimulated on DRS. The 2015 ATA guideline suggests the use of the Tg levels for defining BIR on DRS in PTC patients who underwent RAI ablation: stimulated Tg >10 ng/mL or unstimulated Tg >1 ng/mL. Subsequently, several previous studies have suggested the following cut-off points for defining BIR in PTC patients who have not undergone RAI ablation treatment: stimulated Tg >10 ng/mL or unstimulated Tg >5 ng/mL.[14,15] However, this definition has not been fully validated, and our study results should be interpreted with caution, as patients who had not undergone RAI ablation were included in the training (64.6%) and validation (31.4%) cohorts.

The prognoses of patients with BIR on DRS are generally favorable; 56–68% of patients with BIR are finally classified to have "no evidence of the disease," 19–27% experience a persistent abnormal Tg value without structural recurrence, and 8–17% develop structural recurrence over a 5–10-year follow-up period.[12,13,16] Interestingly,

previous studies have not reported the disease-specific death rates in patients with BIR over a 10-year follow-up period, implying that the clinical significance of BIR in disease-specific mortality is relatively low. However, the clinical impact of BIR cannot be ignored concerning the disease recurrence, which may lead clinicians to drive excessive efforts toward the identification and treatment of potential recurrence. Tuttle et al reported that patients with BIR on DRS are exposed to subsequent aggressive TSH suppression, frequent imaging, and additional treatments, including empirical RAI ablation or surgical exploration.[11] In these patient, surgical exploration may be associated with complications, including parathyroid gland and recurrent laryngeal nerve injury.[17–19] Therefore, along with efforts aimed at reductions in the excessive management of patients with BIR, clinicians should consider the possibility of frequent evaluation and active treatment strategy for the identification and treatment of potential recurrence in patients with a high predicted risk of BIR on DRS. Our prediction model may contribute to BIR risk prediction in PTC patients and aid in the formulation of reliable and individualized management strategies.

Risk factors for BIR on DRS in PTC patients have been investigated previously. In a study on 554 PTC patients,

Danilovic et al reported that the presence of gross ETE was a significant risk factor for BIR,[7] whereas Mendoza et al demonstrated the following risk factors: male sex, extent of LN metastasis, ETE, and multifocality.[20] The total RAI treatment dose and presence of *BRAF V600E* mutations have also been identified as significant risk factors for BIR in PTC patients.[4,21] However, these studies only reported the possible risk factors for BIR and have not suggested a comprehensive prediction model for calculating the BIR risk in individual PTC patients. The risk factors identified in our study correspond to those noted in previous studies, and our model can predict patients' risk for BIR on DRS considering the relative weight of individual risk factors. Although the clinicopathologic characteristics noted in the training and validation cohorts were different, the consistent performance of the prediction model in the two cohorts implied its general applicability.

Our study had several limitations. First, the *BRAF V600E* mutation status of patients was not evaluated, which was associated with poor prognoses in PTC.[22,23] Although Kowalska et al reported that the *BRAF V600E* mutation status was not significantly correlated with the response to initial therapy,[24] the inclusion of *BRAF* mutation status as a predictor may enhance the performance of our prediction model. However, we were unable to include this parameter because a large number of patients had missing data. Second, data were retrospectively collected, indicating a potential selection bias. Third, the number of BIR events in our study was relatively small, which may have negatively affected the power of the prediction model. Fourth, although the size of the LN metastasis was considered an important risk factor for recurrence in the 2015 ATA guideline, this was not investigated in our study. Finally, our study only used the serum Tg levels and not the anti-TgAb levels for defining BIR on DRS, which was inconsistent with the definition of BIR in the 2015 ATA guideline. However, we employed the ambiguous cut-off range for the anti-TgAb levels to define BIR on DRS, which was only referred to as the "rising level" in the 2015 ATA guideline. Although previous studies have reported on the association between the increasing anti-TgAb levels and PTC recurrence,[25–29] they have not suggested a distinct cut-off range for increasing the anti-TgAb values. Despite these several limitations, our study established and validated the nomogram with good performance for predicting BIR risk on DRS in PTC patients. A more properly designed future study would help improve the reliability of our prediction model.

Conclusion

In conclusion, this study presented a nomogram that could effectively predict BIR risk in the DRS of PTC patients without structural recurrence. Further studies are needed to evaluate the practicality of our nomogram in real-world clinical situations.

Ethics Approval and Informed Consent

This study was approved by the Institutional Review Board (IRB) of St. Vincent's Hospital (IRB No. XC20RADE0141). All procedures involving human participants were in accordance with the ethical standards of the institutional research committee and the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All participants provided written informed consent for storing medical information in the hospital database and use of this information for research purposes.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas, took part in drafting, revising or critically reviewing the article, gave final approval of the version to be published, have agreed on the journal to which the article has been submitted, and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors have no conflicts of interest for this work to declare.

References

1. Lu Y, Jiang L, Chen C, Chen H, Yao Q. Clinicopathologic characteristics and outcomes of papillary thyroid carcinoma in younger patients. *Medicine*. 2020;9915:e19795. doi:10.1097/MD.0000000000019795

2. Lang BH, Shek TW, Wan KY. Impact of microscopic extra-nodal extension (ENE) on locoregional recurrence following curative surgery for papillary thyroid carcinoma. *J Surg Oncol*. 2016;113(5):526–531. doi:10.1002/jso.24180

3. Aldawish M, Jha N, McEwan AJ, Severin D, Ghosh S, Morrish DW. Low but measurable stimulated serum thyroglobulin levels <2 μg/L frequently predict incomplete response in differentiated thyroid cancer patients. *Endocr Res*. 2014;39(4):157–163.

4. Deng Y, Zhu G, Ouyang W, et al. Size of the largest metastatic focus to the lymph node is associated with incomplete response of PN1 papillary thyroid carcinoma. *Endocr Pract*. 2019;25(9):887–898. doi:10.4158/EP-2018-0583

5. Shen FC, Hsieh CJ, Huang IC, Chang YH, Wang PW. Dynamic risk estimates of outcome in Chinese patients with well-differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation. *Thyroid*. 2017;27(4):531–536. doi:10.1089/thy.2016.0479

6. Qu Y, Huang R, Li L. Low- and high-dose radioiodine therapy for low-/intermediate-risk differentiated thyroid cancer: a preliminary clinical trial. *Ann Nucl Med*. 2017;31(1):71–83. doi:10.1007/s12149-016-1133-4

7. Danilovic DLS, Castroneves LA, Suemoto CK, et al. Is there a difference between minimal and gross extension into the strap muscles for the risk of recurrence in papillary thyroid carcinomas. *Thyroid*. 2020;30(7):1008–1016. doi:10.1089/thy.2019.0753

8. Yoon J, Yoon JH, Han K, et al. Ultrasonography surveillance in papillary thyroid carcinoma patients after total thyroidectomy according to dynamic risk stratification. *Endocrine*. 2020;69(2):347–357. doi:10.1007/s12020-020-02347-9

9. Randolph GE, Duh QY, Heller KS, et al. The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. *Thyroid*. 2012;22(11):1144–1152. doi:10.1089/thy.2012.0043

10. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26(1):1–133.

11. Tuttle RM. Optimal management of a biochemical incomplete response to therapy in differentiated thyroid cancer: aggressive treatment or cautious observation. *Endocrine*. 2014;46(3):363–364. doi:10.1007/s12020-014-0213-2

12. Tuttle RM, Tala H, Shah J, et al. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. *Thyroid*. 2010;20(12):1341–1349.

13. Vaisman F, Momesso D, Bulzico DA, et al. Spontaneous remission in thyroid cancer patients after biochemical incomplete response to initial therapy. *Clin Endocrinol*. 2012;77(1):132–138. doi:10.1111/j.1365-2265.2012.04342.x

14. Momesso DP, Tuttle RM. Update on differentiated thyroid cancer staging. *Endocrinol Metab Clin North Am*. 2014;43:401–421. doi:10.1016/j.ecl.2014.02.010

15. Momesso DP, Vaisman F, Yang SP, et al. Dynamic risk stratification in patients with differentiated thyroid cancer treated without radioactive iodine. *J Clin Endocrinol Metab*. 2016;101(7):2692–2700. doi:10.1210/jc.2015-4290

16. Vaisman F, Tala H, Grewal R, Tuttle RM. In differentiated thyroid cancer, an incomplete structural response to therapy is associated with significantly worse clinical outcomes than only an incomplete thyroglobulin response. *Thyroid*. 2011;21(12):1317–1322. doi:10.1089/thy.2011.0232

17. Calò PG, Medas F, Conzo G, et al. Intraoperative neuromonitoring in thyroid surgery: is the two staged thyroidectomy justified? *Int J Surg*. 2017;41(Suppl 1):213–S20. doi:10.1016/j.ijsu.2017.02.001

18. Conzo G, Docimo G, Mauriello C, et al. The current status of lymph node dissection in the treatment of papillary thyroid cancer. A literature review. *Clin Ter*. 2013;164(4):2242–2246.

19. Docimo G, Tolone S, Ruggiero R, et al. Total thyroidectomy without prophylactic central neck dissection combined with routine oral calcium and vitamin D supplements: is it a good option to achieve a low recurrence rate avoiding hypocalcemia? A retrospective study. *Minerva Chir*. 2013;68(3):321–328.

20. Mendoza ES, Lopez AA, Valdez VA, et al. Predictors of incomplete response to therapy among filipino patients with papillary thyroid cancer in a tertiary hospital. *J Endocrinol Invest*. 2016;39(1):55–62. doi:10.1007/s40618-015-0319-2

21. Zern NK, Clifton-Bligh R, Gill AJ, et al. Disease progression in papillary thyroid cancer with biochemical incomplete response to initial therapy. *Ann Surg Oncol*. 2017;24(9):2611–2616. doi:10.1245/s10434-017-5911-6

22. Xing M, Liu R, Liu X, et al. BRAF V600E and TERT promoter mutations cooperatively identify the most aggressive papillary thyroid cancer with highest recurrence. *J Clin Oncol*. 2014;32(25):2718–2726. doi:10.1200/JCO.2014.55.5094

23. Xing M, Westra WH, Tufano RP, et al. BRAF mutation predicts a poorer clinical prognosis for papillary thyroid cancer. *J Clin Endocrinol Metab*. 2005;90(12):6373–6379. doi:10.1210/jc.2005-0987

24. Kowalska A, Walczyk A, Kowalik A, et al. Response to therapy of papillary thyroid cancer of known BRAF status. *Clin Endocrinol*. 2017;87(6):815–824. doi:10.1111/cen.13423

25. Spencer CA, Takeuchi M, Kazarosyan M, et al. Serum thyroglobulin autoantibodies: prevalence, influence on serum thyroglobulin measurement, and prognostic significance in patients with differentiated thyroid carcinoma. *J Clin Endocrinol Metab*. 1998;83(4):1121–1127.

26. Chung JK, Park YJ, Kim TY, et al. Clinical significance of elevated level of serum antithyroglobulin antibody in patients with differentiated thyroid cancer after thyroid ablation. *Clin Endocrinol*. 2002;57(2):215–221. doi:10.1046/j.1365-2265.2002.01592.x

27. Görges R, Maniecki M, Jentzen W, et al. Development and clinical impact of thyroglobulin antibodies in patients with differentiated thyroid carcinoma during the first 3 years after thyroidectomy. *Eur J Endocrinol*. 2005;153(1):49–55. doi:10.1530/eje.1.01940

28. Seo JH, Lee SW, Ahn BC, Lee J. Recurrence detection in differentiated thyroid cancer patients with elevated serum level of antithyroglobulin antibody: special emphasis on using (18) F-FDGPET/CT. *Clin Endocrinol*. 2010;72(4):558–563. doi:10.1111/j.1365-2265.2009.03693.x

29. Adil A, Jafri RA, Waqar A, et al. Frequency and clinical importance of anti-Tg auto-antibodies (ATG). *J Coll Physicians Surg Pak*. 2003;13(9):504–506.